

# 携帯端末を用いたミーティング定量評価システムの構築

宇佐美 格<sup>1,a)</sup> 王 亜楠<sup>2,b)</sup> 高橋 淳二<sup>3,c)</sup> 齊藤 裕樹<sup>4,d)</sup> 戸辺 義人<sup>3,e)</sup>

受付日 2015年6月1日, 採録日 2015年11月5日

**概要:** グループにおける意思決定手段としてミーティングは重要であり, 運営者には活発な議論を行う議事進行が求められる. 運営者は実施されたミーティングに対して適切な評価を行い, ミーティングにフィードバックを得られることが必要である. しかし, 運営者がミーティングの音声ログや会話ログからミーティングの評価を行う場合, 評価指標があいまいであり定量的な評価が得られない. また, 従来, ミーティングのアクティビティをセンシングによって解析する試みがいくつか存在するが, 定性的評価に限定される, 適用範囲に限られるといった問題がある. 本研究では発言の頻度と遷移に基づくミーティングの活発度を表す指標を定義し, ミーティングの特徴を定量的に評価するシステム KAIHUI の設計と実装を行った. 評価実験で各指標の特性を明らかにし, ユーザによる主観評価との比較実験により計測手法の有効性を確認した.

**キーワード:** ミーティング分析, 発話解析, スマートフォンセンシング

## Quantitative Evaluation System of Meeting Using a Mobile Device

ITARU USAMI<sup>1,a)</sup> YANAN WANG<sup>2,b)</sup> JUNJI TAKAHASHI<sup>3,c)</sup> HIROKI SAITO<sup>4,d)</sup> YOSHITO TOBE<sup>3,e)</sup>

Received: June 1, 2015, Accepted: November 5, 2015

**Abstract:** Meetings are considered to be indispensable for decision making of groups, and thus argument activation by a chairman is required. A chairman needs to evaluate appropriately to a meeting and feed back to that. However evaluation of meetings can be qualitative and obscure if we utilize only recorded conversations. In this paper, to reveal efficiency of meetings, we define three metrics for meetings based on conversational frequency and transition, and we design and implement a system called KAIHUI to evaluate the metrics. Then, we confirm that the measurement study is efficient through the experiments.

**Keywords:** meeting analysis, speech recognition, smartphone sensing

<sup>1</sup> 青山学院大学大学院理工学研究科理工学専攻知能情報コース  
Intelligence and Information, Graduate School of Science  
and Engineering, Aoyama Gakuin University, Sagami-hara,  
Kanagawa 252-5238, Japan

<sup>2</sup> 電気通信大学大学院情報システム学研究科社会知能情報学専攻  
Department of Social Intelligence and Informatics, Graduate  
School of Information Systems, The University of Electro-  
Communications, Chofu, Tokyo 182-8585, Japan

<sup>3</sup> 青山学院大学理工学部情報テクノロジー学科  
Department of Integrated Information Technology, College  
of Science and Technology, Aoyama Gakuin University,  
Sagami-hara, Kanagawa 252-5238, Japan

<sup>4</sup> 明治大学総合数理学部先端メディアサイエンス学科  
Department of Frontier Media Science, School of Interdis-  
ciplinary Mathematical Sciences, Meiji University, Nakano,  
Tokyo 164-8525, Japan

a) itaru@rcl-aoyama.jp

b) wang@ni.is.uec.ac.jp

c) takahashi@it.aoyama.ac.jp

d) hrksaito@meiji.ac.jp

e) yoshito-tobe@rcl-aoyama.jp

## 1. はじめに

近年, インターネットの普及とともに, コミュニケーションの多様化が進んでいる. 従来のコミュニケーション形態である通話や電子メールだけでなく, ビデオ会議や SNS (Social Network Service) といった新たなコミュニケーションサービスが登場している. しかし, 今なお, 対面でのミーティングは有効なコミュニケーション手段である. たとえば, ブレーンストーミングのように連続して多彩なアイデアを出し続け, 円滑な会話の流れを必要とする場面では, 対面でのミーティングはきわめて有効である.

対面でのミーティングは様々な場面で用いられ, 効率化することは重要な問題である. 従来の効率性を改善するシステムには, ミーティングの様子を議事録, 音声ログや映像ログにし, 後から参照できるミーティング支援システム

がある。しかし、ミーティングの客観的な指標を用いずに、ログ情報を参照する定性的評価だけでは問題点を明確にすることが難しい、ミーティングを定量化指標によって評価、比較することができれば、ミーティングの改善に向けた情報として利用することが可能である。

本研究では、ミーティングの音声情報をセンシングし、発話者解析を行った後、発話の時間長と頻度、発話者の遷移確率をもとにミーティングの活発度を定量的に評価する手法を提案する。具体的には、参加者らのインタラクションに基づいて、参加者らが平等に発言しているかどうかを表す公平度、ミーティング全体をリードする割合を表す支配度、参加者の意見を導き出すファシリテーションを表す調停度の3つの指標を用いて、ミーティングの活発度を定量化する。さらに、本指標の計算方法に基づき、スマートフォンを用いてミーティング中の音声記録から各指標を算出するシステム KAIHUI を構築した。KAIHUI では、ミーティングの発話者を識別するために指向性マイクといったような特殊なデバイスに依存せず、一般に普及しているスマートフォンのマイクを用いることが特徴である。KAIHUI により得られたミーティング活発度の指標により、異なるミーティングの活発度を比較し分析を行うことや、活発なミーティングを促す助言といった用途への応用が期待できる。

本論文では以下、2章では、既存研究について述べ本研究の特徴や位置づけを明らかにする。3章では、活発度指標の定義について詳しく述べる。4章では、KAIHUI の機能要求の分析を行いシステムの設計と実装を行う。5章では、KAIHUI の評価実験の内容と結果について述べる。6章では、本論文のまとめを述べる。

## 2. 関連研究

これまで、ミーティングの様子を議事録、音声、映像などによって記録し、参照可能な様々なシステムが開発されてきた。たとえば、Nagao は、ミーティングの音声、映像が含まれる議論内容にメタデータを付したコンテンツとして利用可能にし、ミーティングの参加者らは日付や参加者の情報から議事内容を検索するだけでなく、現在進行中の議論と類似の議論を過去の議論コンテンツから探し出したり、ある質問に対する回答を議事録の内容に基づいて生成したりすることを可能とした [1]。しかし、あらゆる話題に対応する知識データベースの構築が必要となり、会話を完全に理解するのは困難である。

ミーティングの効率を上げるためのガイドラインが多く存在する。しかし、普遍的なガイドラインがなく、定性的な指標が多いため、具体的なミーティングの評価は困難である。これに対して、Jonker らはミーティングの内容を詳しく解析するために、ミーティングのドメイン情報に基づいたミーティングプロトコル解析の手法を提案している [2]。

この研究では、発話内容を分析したうえでミーティングのフローを定義づけしようとしているが、記録したミーティング音声をオフラインで分析する必要があるため本研究で目指すようなリアルタイム処理には不向きである。本研究では、発話内容には踏み込まず、ミーティング参加者の発話の有無のみに着目することにより、モバイル機器利用での簡易な定量化を目指す。

Kulyk らは少人数向けのミーティングにおける非言語的な振舞いをリアルタイムでフィードバックするシステムを提案している [3]。このシステムでは参加者の発話時間や注目される振舞いをリアルタイムで視覚化することにより、効果的なミーティングが得られることを示している。しかし、リアルタイムで机の上に設置されたプロジェクタからフィードバック情報を提示する必要があるため、限られた場所でしか使えず、導入するのにコストを要するという問題がある。

対面的コミュニケーションの場においては、様々な情報伝達的手段が存在する。情報伝達には、言葉だけでなく、視線やジェスチャといった非言語的な要素が含まれる。そこで、中田らは、会話における非言語情報の出現パターンをデータマイニングし、三者対面時の発話と視線の関係や発話者と聞き手による指さしの使われ方に関するパターンが自動抽出できることを示した [4]。しかし、会話構造には多くの非言語行動が関わっており、より多くのパターンの抽出は容易ではない。本研究では、ミーティングの音声情報のみ注目し、参加者らのインタラクションを分析することで、ミーティング活発度の定量化を試みる。

守屋らは、ミーティング参加者の音声情報に基づく会話活性度の推定方法および自然会話における活性度の自動推定の可能性を検討している [5]。ミーティングの会話構造を反映するオーバーラップ率・平均発話潜時・発話率の3つの指標を用いて、会話活性度を推定する可能性を示している。また、豊田らは、発話状態時間長に着目した対話雰囲気推定システムを提案している [6]。これらに対し本研究は、話者発話頻度や話者遷移確率に基づくミーティング活発度の定量化指標を定義するものである。

本研究ではミーティング参加者のインタラクションを明確にするために音声解析に基づく話者識別が必要になる。荒木らは、音声区間検出器 (VAD) で検出した音声区間における音声到来方法 (DOV) を分類することで、会議音声の話者識別を行う手法を提案している [7]。しかし、この手法では、ステレオマイク3台を用いることと話者の位置情報を事前に固定することが必要となる。本研究は、より利便性の高いシステムを構築するためスマートフォンのマイクを使用することを前提に、VAD による話者識別ではなく、話者らの声道特性を表す MFCC (Mel-Frequency Cepstrum Coefficients) [8], [9] に基づいた K-means クラスタリング [10] を用いて話者識別を行う。

スマートフォンを用いてオーディオセンシングのような累積処理をする場合には、スマートフォンのバッテリー消費が激しくなる。これに対し、Georgievらは低消費電力コプロセッサの性能を考慮した省電力化を図っている [11]。我々の研究はオーディオセンシングそのものではなく、ミーティングの定量化を目的としており、本研究においてもこの手法を用いることは可能である。

### 3. ミーティング活発度の定量化指標

本システムはミーティングの活発度を定量化することが目的である。ミーティングの改善を目的とした従来研究は、部屋全体を利用するような大規模なシステムを用いることが多い。これらのシステムは導入コストが高く、特定の部屋のみでしか利用できないといった空間的な制約を生じる問題がある。本研究では、これらとは別のアプローチをとり、スマートフォンのマイク機能をセンサとして用い、クライアントであるスマートフォンからサーバに音声情報を送信し計算を行うことでミーティングの活発度を取得する手法を考える。ミーティング参加者がスマートフォンを用いるのみで利用することができるため、簡便で、適用範囲が広く、空間的な制約にもしばられないシステムを考えることができる。

一般に、スマートフォンのマイクは人の音声を入力として想定されているため、人の音声を集音するうえでは問題ないと考えられる。ただし、ノイズの影響は大きく、雑踏の中での音声収集は、信号対雑音比が小さくなり、雑踏など、環境音が大きな中で本研究のシステムを利用するのは難しいと考えられ、本システムは環境音の小さな会議室での利用を想定している。

さらに、定量化された指標を基に、異なるミーティングを比較し差異を分析する、直接的にミーティングの改善を支援するシステムといった応用が考えられる。たとえば、ミーティングに参加している各個人に適切なフィードバックを返すようなシステムや、リアルタイムでミーティングの質を高めるためのシステムなどが考えられる。

#### 3.1 システムの流れ

KAIHUIはクライアントサーバモデルで構成される。まずクライアントのスマートフォンを用いて発言を録音し、サーバに音声データを送信する。音声データの解析はサーバ上で行われ、結果がクライアントに返送される。

まず、音声データから話者を識別するための前処理としてMFCC (Mel-Frequency Cepstrum Coefficients, メル周波数ケプストラム係数) を求める。MFCCは、人の周波数知覚特性を考慮した重みづけ特徴量により人の声道情報を表すものである。本システムでは、参加者の音声情報は12次元のMFCCベクトルで表される。

次に、参加者らの発話時間・順番・回数などの情報を得

るためクラスタリングにより発話者を識別するプロセスを行う。従来の話者識別には、教師あり学習であるGMM (Gaussian Mixture Model) [12] アルゴリズムが用いられることが多いが、事前に参加者の学習データを用意する必要があるため、その場限りのミーティングに対応できないという問題がある。そこで、本研究ではミーティング参加者を特定しないことを前提に、K-means法での参加者クラスタリングを用いる。K-means法を用いたクラスタリングの結果は初期値に依存するという問題 [13] があるため、ミーティングの初めの各参加者の発言を初期値とする。次節では、クラスタリングによって発話者が識別された後、ミーティング活発度を示す各指標値の計算方法を示す。

#### 3.2 活発度の定量化指標

ミーティングという言葉で表現できる実体は、参加人数、目的、詳細アジェンダの有無などに応じて、様々な形態が存在する。たとえば、発注側と受注側で納入システムの仕様打ち合わせもミーティングであるし、会社組織において会社の方針が50名ほどの社員に伝達される場合もミーティングと呼ぶことが可能である。本研究で対象とするミーティングは、その中でも、3m以内の同一空間にいる3~10名の参加者が、相互に意見を出し合う形態を想定しており、アジェンダが決まっていた短時間のうちに数々の議題をこなしていく形式でもよいし、ブレインストーミングとして自由にアイデアを出すような形式であってもよい。なお、3m以内というのは目安であり、1台のスマートフォンで音声を取得できる範囲であればよい。

活発なミーティングとは、あるテーマに対して多くの参加者が積極的に議論すること、ミーティングをリードする人がいること、ミーティングのリーダは参加者の意見を平等に引き出していることの3要素であると定義する。議論するにあたり、発言が公平に行われているか否かを評価するには、参加者の発言時間と発言回数の2種類を考慮する必要がある。今回、発言時間に着目した評価を行うのに議論公平度、発言回数に着目した評価を行うのに議論支配度、各参加者の平等な発言機会に着目した評価を行うのに議論調停度という指標を設けることとした。以下、ミーティング活発度を測る3つの指標を示す。

##### 3.2.1 指標1：議論公平度

発言時間が長い参加者は議論への参加度合いが高いと考えられる。活発なミーティングには、参加者が公平に議論に参加していることが条件となる。ミーティング中の発言時間に注目し、参加者全員が平等に議論を行っているかを示す指標として議論公平度を定義する。

ここで、まず発言時間割合を定義する。発言時間割合  $P_i^E$  とは、以下の式 (1) のように、各参加者の発言時間  $t_i$  ( $i = 1, 2, \dots, K$ ) ( $K$  は参加者数) を発言時間の総和で割ったものとなる。

$$P_i^E = \frac{t_i}{\sum_{i=1}^K t_i} \quad (1)$$

次に、議論公平度を、Shannonのエントロピーの性質 [14] を用いて定義する。参加者全員が同じ発言確率である場合は議論公平度が最大値になり、発言者は1名で残りの参加者が言葉を発しない場合は議論公平度が0になる。さらに、参加者数によってエントロピーの最大値が異なるため、人数による最大エントロピーの正規化を行う。以上から、議論公平度 (Equity) を式 (2), (3) のように定義する。

$$H_E = - \sum_{i=1}^K P_i^E \log_2 P_i^E \quad (2)$$

$$\text{Equity} = \frac{H_E}{\max_K H_E} \quad (3)$$

### 3.2.2 指標 2：議論支配度

ある参加者の発言頻度が高い場合、この参加者をミーティングの支配者であるとする。ここでは、議論を主導する人の指標として、支配者の数を正規化したものを議論支配度と定義する。

各参加者の発言回数を  $q_i$  ( $i = 1, 2, \dots, K$ ) とし、ある参加者の発言回数を発言回数の総和で割ったものを発言頻度とする。発言頻度は式 (4) で得られる。

$$P_i^C = \frac{q_i}{\sum_{i=1}^K q_i} \quad (4)$$

次に、この発言頻度が一定以上の参加者を支配者とする。ここで、支配者を識別するしきい値は、ミーティングの質により異なると考えられる。たとえば、ブレインストーミングのように各参加者が平等に発言するケースでは、平均以上の頻度で発言した参加者を支配者と見なせばよいが、複数の議題や意見をまとめながら司会者が進行するミーティングでは、司会者はミーティングの進行が役割であり、実質的には議論の中身を左右するのは司会者の制御のもとで発言する各参加者であることもありうる。このようなケースでは、しきい値を下げ他の参加者の発言状況を加味する必要があると考えられる。

参加者  $i$  が支配者であるかどうか ( $N(i)$ )、支配者の人数 (Controller), 議論支配度 (Domination) は、決められたしきい値  $\text{threshold}$  を基に、以下の式 (5), (6), (7) により求められる。

$$\text{Controller} = \sum_{i \in K} N(i) \quad (5)$$

$$\text{where } N(i) = \begin{cases} 1 & \text{if } P_i^C \geq \text{threshold} \\ 0 & \text{otherwise.} \end{cases} \quad (6)$$

$$\text{Domination} = \frac{\text{Controller}}{K} \quad (7)$$

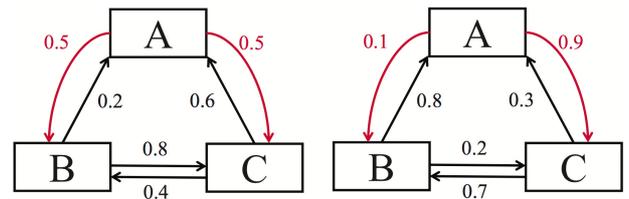


図 1 発話者の状態遷移

Fig. 1 State transition of speakers.

### 3.2.3 指標 3：議論調停度

活発な議論を促すためには、司会進行役である支配者の役割が重要である。支配者は意見を引き出すために、発言権を公平に与える必要がある。ここでは、支配者による発言権が平等に与えられているかどうかを表す議論調停度を定義する。

発言者の遷移が単純マルコフ連鎖であることを想定し、支配者  $i$  における遷移確率を  $P_{ij}^M$  ( $i \neq j$ ) とする。図 1 は各参加者の発言の遷移確率を示している。図の左側では、参加者 A から参加者 B と参加者 C への遷移確率は等しいことから、参加者 A は双方に対して平等に発言機会を与えていることが分かる。図の右側は参加者 A から参加者 B への遷移確率が 0.1、参加者 C への遷移確率が 0.9 であることから、参加者 C の発言によるバイアスが加わり参加者 B の主張を述べる機会が少ない状況が分かる。

ここでは、遷移確率によるエントロピーを求めることで、議論調停度を算出する。議論調停度 (Mediation) は、以下の式 (8), (9) で表される。

$$H_{Mj} = - \sum_{i=1, i \neq j}^K P_{ij}^M \log_2 P_{ij}^M \quad (8)$$

$$\text{Mediation}_j = \frac{H_{Mj}}{\max_K H_M} \quad (9)$$

ミーティングに複数の支配者が存在する場合、式 (10) のように議論調停度は支配者の平均調停度となる。

$$\text{Mediation} = \frac{\sum_{\forall j \in \{1, 2, \dots, K\}: P_j^C \geq \text{threshold}} \text{Mediation}_j}{\sum_{\forall j \in \{1, 2, \dots, K\}: P_j^C \geq \text{threshold}} N(j)} \quad (10)$$

## 4. KAIHUI の設計と実装

KAIHUI はミーティング参加者らの発言時間、発言頻度や発言順番といったインタラクション情報を解析し、ミーティング活発度の指標を求めるシステムである。システム構成および機能を分割する各モジュールを図 2 に示す。KAIHUI はクライアントとサーバによって構成される。本研究では、適用範囲の広いシステムを目指すため、クライアントにはスマートフォンを用いる。また、サーバは計算性能が求められるため専用サーバを用いる。以下、各プロセスの手順について述べる。

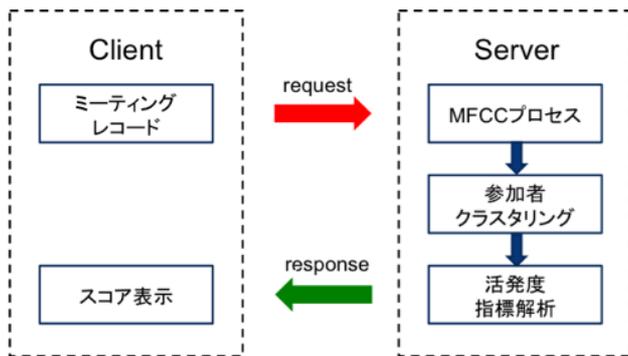


図 2 KAIHUI のシステムモデル  
Fig. 2 System model of KAIHUI.

ミーティングレコード スマートフォンを用いて発言ごとに録音し、音声データをサーバに送る。  
 MFCC プロセス 参加者クラスタリングの前処理として、MFCC に基づく特徴量抽出を行う。  
 参加者クラスタリング K-means 法を用いて、参加者クラスタリングを行う。  
 活発度指標解析 参加者クラスタリングの結果に基づき、ミーティング活発度の指標解析を行う。  
 スコア表示 3つの指標値をクライアントに送信し、ユーザに提示する。

4.1 KAIHUI の実装

以下、システム KAIHUI の実装について述べる。前述のとおり、本システムはスマートフォンのクライアントとサーバによって構成される。

4.1.1 クライアント側の実装

クライアント側ソフトウェアは、Android 上で動作するアプリケーションとして実装した。1 台の Android 端末がミーティングの参加者全員の発言の収集を行う。本システムでは、あらかじめ参加者数と K-means 法クラスタリングの初期値が必要となる。そこで、クライアントはミーティングの音声全体を録音しサーバに送信するのではなく、参加者の発言ごとにデータを分割して録音する機能を有する。なお、発言単位に音声を区切るためには、発言間の無音区間により切り出しを行う手法や、発言者が交代した際に切り出しを行う手法などが考えられる。また、複数の発言者がオーバーラップした際にはさらに独立成分分析を用いた手法 [15] などにより 1 つの音声データから複数の発言者を識別する手法などを検討する必要がある。本研究では、発言者が発言を行う際に明示的にクライアント上でボタンを押してから発言するシンプルで確実な手法を用い、研究の第 1 段階として発言の頻度と時間に注目した分析を正確に行うことを目的とした。まず、ミーティング開始時に参加人数を入力し冒頭で全参加者が順番に自己紹介などを行い各参加者の最初の発言を初期値とする。各参加者の発言の音声データ (1.wav, 2.wav, ..., N.wav) はサーバに送信

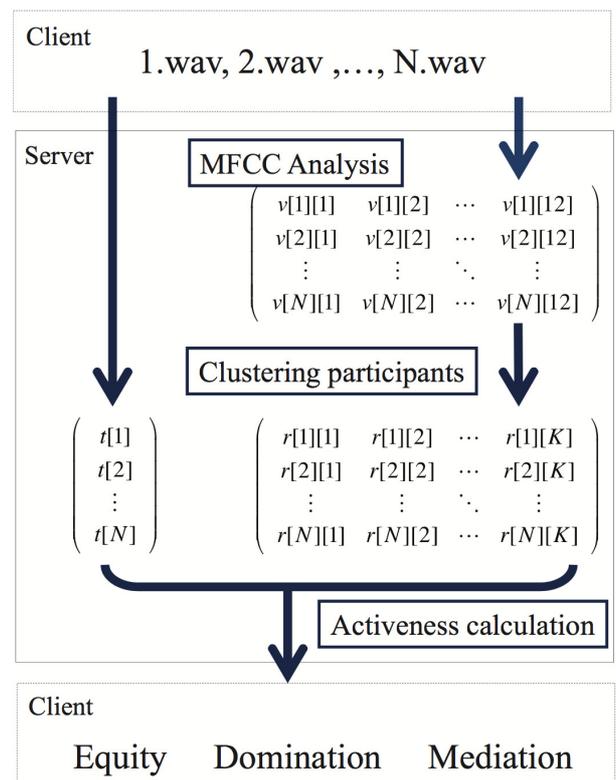


図 3 サーバ側の音声処理プロセス  
Fig. 3 Sound processing in server.

され、参加者数とクラスタリング時の初期値を得る。

スマートフォンではプロセッサ性能が限られるため、MFCC プロセスなどの計算量の多い処理はサーバで行い、サーバで計算された解析結果をクライアントに表示する構成とした。

4.1.2 サーバ側の実装

サーバの実装の詳細を図 3 に示す。サーバでは、 $N$  個のファイル (1.wav, 2.wav, ..., N.wav) の発言データをクライアントから受け取り、MFCC 分析によって音声処理を行う。結果、 $t[1] \dots t[N]$  を成分を持つ  $N$  次元の時間ベクトル  $\mathbf{t}$ ,  $v[1][1] \dots v[N][12]$  を成分を持つ  $N \times 12$  の MFCC 行列  $\mathbf{V}$  が生成される。参加者クラスタリングの処理においては、ミーティング冒頭で全参加者 ( $K$  名) の自己紹介などの発言の MFCC 行列  $v[1][1] \dots v[K][12]$  を初期値とし、K-means 法を用いたクラスタリングを行う。その結果、 $r[1][1] \dots r[N][K]$  を成分を持つ  $N \times K$  の 2 値行列  $\mathbf{R}$  が得られる。なお、 $N$  は発言数、 $K$  は参加者数である。この 2 値行列は各成分が 0 か 1 かを表す行列であり、各参加者の発言頻度、発言順番といったインタラクション情報が含まれる。活発度指標解析は、時刻ベクトル  $\mathbf{t}$ , 2 値行列  $\mathbf{R}$  を用い 3 章で述べた指標が解析される。

なお、このクラスタリング処理は、ミーティングが終了しすべての発言に対して行う方法もあるが、ミーティング中のあるセッションや、ミーティング途中のある時点から一定時間遡ったウィンドウを単位として指定された時間区

間にある発言群を切り出すことでリアルタイムに分析を行うことも可能である。

## 5. 評価実験

KAIHUI システムを用いて、6 種類のミーティング (A, B, C, D, E, F) の活発度を計測した。ミーティング活発度指標の有効性を評価し、さらにユーザによる主観評価との比較を行う。

実験するにあたり、すべての試行を中央に机が置かれた同一の会議室で行い、スマートフォンはどの参加者とも 1m 以内の中央に配置し、参加者間の距離は最大で 2m 以内になる距離で行った。

### 5.1 実験対象のミーティング

すべてのミーティングについて被験者は互いに知り合い同士の 20 歳代の学生であった。また性別については、ミーティング C は女性のためのミーティングで、それ以外は男性のためのミーティングであった。

各ミーティングのテーマおよび参加者の各テーマに対する前提知識、議論の進行は以下のとおりであった。

**ミーティング A【為替レートに関するテーマ】** 1 名が為替に関する疑問を投げかけ、それについて詳しいもう 1 名が答える流れを中心に議論を進行していった。

**ミーティング B【興味のある研究室に関するテーマ】** これから配属研究室を選択する学生 3 名が特に関心のある研究室についての議題を投げかけ、上級生 1 名が議論を主導した。

**ミーティング C【テレビドラマに関するテーマ】** 参加者全員が前提知識を有し興味が共通するドラマについて、全員が同じ立場で種々の議論を行った。

**ミーティング D【スマートフォンのゲームに関するテーマ】** 当該分野に精通する 2 名と、そうでない 2 名が議論を行った。

**ミーティング E【研究内容に関するテーマ】** ある研究テーマに取り組んでいる学生 1 名と、そのサポート役の大学院生 1 名が主導的に議論を進めた。当ミーティングでは、この 2 名が主導するように指示を行った。その他の 2 名は必要に応じて受け答えを行った。

**ミーティング F【仕事の姿勢に関するテーマ】** 1 名の学生に司会者となるように指示を与え、社会人としての仕事の姿勢について解説するように議論を進めた。その他の 3 名は必要に応じて受け答えを行った。

### 5.2 KAIHUI システムの評価

クライアントの実験環境を表 1 に、サーバの実験環境を表 2 にそれぞれ示す。この環境下で、4 名でのミーティングを行い、各指標の精度について評価を行った。

まず、クラスタリング精度を表 3 に示す。K-means 法

表 1 クライアント環境

Table 1 Specification of client.

Device	Samsung Galaxy S4
CPU	Quad-core/1.9 GHz
OS	Android 4.4.2
Recording	Mono-16 bit Sample rate 16 KHz

表 2 サーバ環境

Table 2 Specification of server.

CPU	Intel(R) Xeon(R) CPU 2.66 GHz
Memory	4 GB
OS	Linux Red Hat 4.4.7-4
Language	Python 2.7

表 3 参加者クラスタリングの識別率

Table 3 Recognition rate of clustering for participants.

	ミーティング 発言時間 (s)	ミーティング 発言回数	参加者クラスタリング 識別率
A	738	27	63%
B	498	54	100%
C	405	79	65%
D	183	38	72%
E	272	23	78%
F	288	10	100%

クラスタリングでは、基本的にすべてのデータをいずれかのクラスタに所属させ、各クラスタの中心および各データの中心からの距離に基づきデータの再所属を反復的に行う手法をとるため、すべての発言はいずれかの話者に識別される。ここでの参加者クラスタリング識別率とは、すべての発言数に対する、クラスタリング結果が正解と一致していた発言数の割合である。参加者クラスタリングの識別率は最小 63% から最大 100%、平均で 81% となった。また、特にミーティング A, C は精度が低い結果となった。

### 5.3 ミーティング活発度指標の評価

ミーティング活発度の評価実験に先立ち、前節の参加者クラスタリング識別率が低いミーティングに関して原因を検討した結果、誤認識理由として、元々の声質が似ているためクラスタどうしのクラスタ中心の距離が十分ではない、他のデータにはない音声データがあるためデータ（発言）からクラスタ中心までの距離が長い、発言者がどちらともつかないため複数のクラスタ中心からデータまでの距離がほぼ等しいといったことが考えられた。

このように、データが 1 つのクラスタへの帰属率が低い場合、発話者不明と見なし次の計算処理をスキップする、発話認識テストをしスマートフォンの位置を微調整する、確証の度合いを示す指標を加味しユーザに提示する、といった方法が考えられる。以上のような処理を行い識別率

表 4 評価実験結果

Table 4 Evaluation result.

	公平度 (Equity)	支配度 (Domination)	調停度 (Mediation)
A	87	50	85
B	90	50	71
C	90	25	99
D	83	50	59
E	63	50	10
F	28	25	0

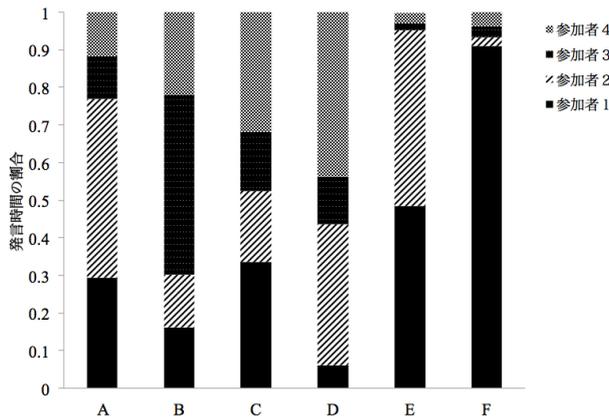


図 4 ミーティング参加者の発言時間割合  
Fig. 4 Ratio of speaking time.

を高めたうえで、発言確率の計算を行うべきであるが、この実験では、参加者間の発言の確率過程の分析に評価の重きを置くため、識別率の低い A, C, D, E については誤認識のあった発言は手作業で修正し、すべての発言者が正しく識別できたという仮定のもとで活発度の評価を行った。6つのミーティングから得られたミーティング活発度を百分率で表したものを表 4 に示す。以下、それぞれの指標の特徴について述べる。

議論公平度について、各ミーティングにおける参加者の発言時間割合を図 4 に示す。議論公平度の高いミーティング A, B, C では、比較的参加者が平等な時間発言している。また、50%を超える時間を占有する発言を行った参加者はいないことや、参加者全員が 10%の時間以上発言していたことも分かる。上記の 3つのミーティングと比較し、議論公平度の低いミーティング E, F では、1名または2名の参加者が 80%を超える時間を占有し、その他の参加者の発言時間は 5%に満たないことが分かる。

議論支配度は発言頻度から計算されるため、各参加者の発言頻度を図 5 に示す。表 4 の支配度は、発言頻度が一定のしきい値以上の値をとる参加者の割合を求めたものである。3.2.2 項で述べたように、ミーティングの性質によってしきい値は異なる値を設定すべきであるが、ここでしきい値を頻度の平均値とすると、4名のミーティングでは 25%以上の頻度で発言した参加者は支配者となる。ミーティング A における参加者 1 と 2 は 25%以上の頻度で発

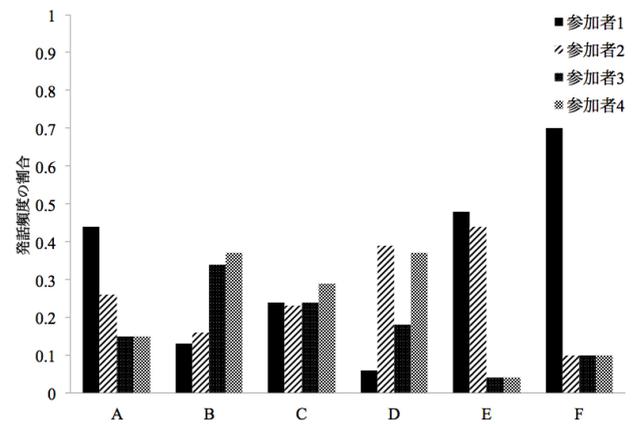


図 5 ミーティング参加者の発言割合  
Fig. 5 Speaking frequency.

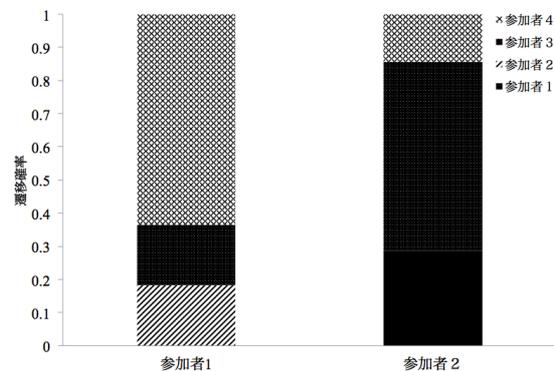


図 6 ミーティング A における参加者 1 と参加者 2 の発言遷移確率

Fig. 6 Probability of state transition from attendee 1 and 2 in meeting A.

言したため、支配者と見なされこのミーティングの支配度は 50 となる。表 4 から、4名のミーティングでは 1名または 2名が支配者となるケースが多いことが分かる。

図 6 は、ミーティング A の支配者である参加者 1, 2 が自らの発言の後どの参加者の発言に遷移したかを表した図である。参加者 1, 2 の遷移を合わせると、支配者ではない参加者 3, 4 への遷移確率はそれぞれ 37.7%, 39.0%であることから、このミーティングでは支配者からその他の参加者に円滑に発言が促されていることが分かる。

#### 5.4 主観評価との比較

A, B, C, D, E, F 6つのミーティングを撮影したビデオを 16名で視聴し、主観的解釈によってミーティングの各指標を評価した。アンケートにより、公平度と調停度はそれぞれの度合いを示す 1~5 の 5段階のスコアの採点、支配度は当該ミーティングを支配していた人数について回答を得た。各指標の平均点を求め、3章で述べた各指標との相関関係を調べた。

主観評価による結果を表 5 に示す。これらの結果と表 4 での提案手法の各指標との相関係数は、公平度が 0.79, 支

表 5 主観評価の実験結果

Table 5 The result of subjective evaluation.

	公平度 (Equity)	支配度 (Domination)	調停度 (Mediation)
A	2.4	2.9	2.8
B	3.9	3.4	3.4
C	3.9	3.3	3.9
D	2.2	3.1	2.8
E	1.3	2.6	2.1
F	1.1	2.3	1.3

配度が 0.24, 調停度が 0.92 であった。

支配度の相関が低い理由として, 3.2.2 項で述べたとおり支配度を識別する適切なしきい値がミーティングの質によって異なるためと考えられる。本実験では, 平均以上の頻度で発言した参加者を支配者と見なすようにしきい値を設定したが, 参加者の意見を集約するようなミーティングでは, 発言頻度の偏りが大きくなり, 少ない頻度で発言した参加者も支配者と見なす必要がある。たとえば, ミーティング F では, 司会者が議論の前提を多く話したことから定量評価によって支配度が小さいという結果を得た。しかし, 主観評価でミーティングを支配している人数を調査した結果, 平均 2.3 (名) と定量評価よりも多くの支配者が存在する結果となった。このことから, 司会者以外の参加者の発言を考慮するためにしきい値を下げたほうが良い場合もありうる。そこで, 適切なしきい値を求めたところ, しきい値を 0.16 に設定すると主観評価との相関が 0.90 と最大になることが分かった。

### 5.5 考察

表 3 のとおりクラスタリングによる話者の識別率は, ミーティングによって結果のばらつきが大きいことが分かる。アルゴリズム上の改良としては, MFCC プロセスの次元数を上げることや事前学習を前提としたクラスタリング手法を用いることが考えられる。一方, 実装上の要因としては, スマートフォンのマイクセンサの質やミーティングを行うときの環境雑音により, 元データの質のばらつきが大きいことが考えられる。また, K-means 法では, 初期値のとり方によって結果が異なることが知られている。KAIHUI では, 各参加者の最初の発言を初期値として用いているため, この発言時間が長いほどクラスタリングの識別精度が向上する。

本研究では, 各指標の基として発言時間と発言頻度の 2 つの指標を用いている。ミーティング A~F の発言時間と発言頻度の相関を求めた結果, 最小 0.53, 最大 0.99, 平均 0.78 となり, 場合によっては時間と頻度に近い傾向がある場合も見受けられるものの, 一般には異なる指標として評価するのが適切であるといえる。したがって, 本研究で提案した指標は互いに相関の少ない独立したものと見なすことができる。

## 6. まとめ

本研究では, 参加者のインタラクションに基づきミーティング活発度を表す, 議論公平度, 議論支配度, 議論調停度の 3 つの指標を提案した。さらに, ミーティングの活発度を測定するシステム KAIHUI の設計・実装を行い, 録音されたミーティングの音声データから, 話者クラスタリングの後, ミーティングの定量化指標を計算して提示することを示した。

今回の指標設定ではいずれかの参加者が発言している状況を確認過程により分析するアプローチをとったが, ミーティング時間全体の中での無音時間と発言時間の比率も活発度の評価に有用であると考えられる。

今後の課題としては, ミーティングの質を考慮した指標の改善があげられる。本研究では, 画一的な計算手法で指標の算出を行っているが, たとえば, アイデアを創出するような意見の発散を重視するミーティングと, 出された意見を集約するミーティングでは, 参加者の発言パターンは異なるものと考えられる。参加者の遷移確率などをあらかじめパターンとして学習することにより, ミーティングの質の差異に適合させることは今後の課題である。

さらに, スマートフォンのローカル環境で計算処理が完結するような計算量の少ないアルゴリズムにシステムを改善することも課題である。また, ミーティングだけでなく, 他の対面コミュニケーションに適応できるシステムを目指し, より活発なコミュニケーションを支援するエージェントにも応用する予定である。

### 参考文献

- [1] Nagao, K.: Discussion Mining: Knowledge Discovery from Semantically Annotated Discussion Content, *LANAI 3609*, pp.158-168 (2007).
- [2] Jonker, C.M., Schut, M., Treur, J. and Yolum, P.: Formal Analysis of Meeting Protocols, *LANI 3415*, pp.114-129 (2005).
- [3] Kulyk, O., Wang, J. and Terken, J.: Real-Time Feedback on Nonverbal Behaviour to Enhance Social Dynamics in Small Group Meetings, *MLMI 2005*, LNCS 3869, pp.150-161 (2006).
- [4] 中田篤志, 角 康之, 西田豊明: 非言語情報の出現パターンによる会話状況の特徴抽出, ヒューマンインタフェース学会研究報告集, Vol.11, No.5, pp.89-96 (2009).
- [5] 守屋悠里英, 田中貴紘, 宮島俊光, 藤田欣也: ボイスチャット中の音声情報に基づく会話活性化度推定方法の検討, ヒューマンインタフェース学会論文誌, Vol.14, No.3, pp.283-292 (2012).
- [6] 豊田 薫, 宮越喜浩, 山西良典, 加藤昇平: 発話状態時間長に着目した対話雰囲気推定, 人工知能学会論文誌, Vol.27, No.2, SP-B, pp.16-21 (2012).
- [7] 荒木章子, 藤本雅清, 石塚健太郎, 澤田 宏, 牧野昭二: 音声区間検出と方向情報を用いた会議音声話者識別システムとその評価, 日本音響学会講演論文集 (2008).
- [8] Sharma, S., Jain, N. and Suwalka, I.: Speech Analysis and Feature Extraction using SCILAB, *IJMCTR*, ISSN:

- 2321-0850, Vol.1, No.1 (2013).
- [9] Md. Hasan, R., Jamil, M., Md. Rabbani G. and Md. Rahman, S.: Speaker Identification Using MEL Frequency Cepstral Coefficients, *ICECE 2004*, pp.28-30 (Dec. 2004).
- [10] Mousa, A.: MareText Independent Speaker Identification based on K-mean Algorithm, *International Journal on Electrical Engineering and Informatics*, Vol.3, No.1, pp.100-108 (2011).
- [11] Georgiev, P., Lane, N.D., Rachuri, K.K. and Mascolo, C.: Co-Processor Support for Continuous Audio Sensing on Smartphones, *ACM SenSys*, pp.295-309 (Nov. 2014).
- [12] Reynolds, D.A. and Rose, R.C.: Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models, *IEEE Trans. Speech and Audio Processing*, Vol.3, No.1, pp.72-83 (1995).
- [13] Steinley, D. and Brusco, M.J.: Initializing K-means Batch Clustering: A Critical Evaluation of Several Techniques, *Journal of Classification*, Vol.24, No.1, pp.99-12 (2007).
- [14] Shannon, C.E.: A Mathematical Theory of Communication, *The Bell System Technical Journal*, Vol.27, pp.379-423, 623-656 (1948).
- [15] 牧野昭二, 荒木章子, 向井 良, 澤田 宏: 独立成分分析に基づくブラインド音源分離, 電子情報通信学会技術研究報告, EA2003-45, Vol.103, No.129, pp.17-24 (2003).



宇佐美 格

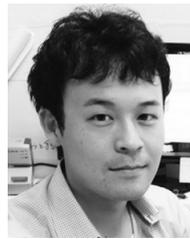
2014年青山学院大学理工学部情報テクノロジー学科卒業。現在、同大学大学院理工学研究科理工学専攻博士前期課程在学中。音声解析, ユビキタスコンピューティングの分野に興味を持つ。



王 亜楠 (学生会員)

2015年青山学院大学理工学部情報テクノロジー学科卒業。現在、電気通信大学大学院情報システム研究科社会知能情報学専攻在学中。センサネットワーク, 人工知能, コンピュータビジョン, サイネージシステムの研究に

従事。



高橋 淳二

2010年名古屋大学大学院工学研究科マイクロ・ナノシステム工学専攻博士後期課程修了。博士(工学)。2010年筑波大学GCOE(サイバニクス)研究員, 2013年名古屋大学大学院工学研究科研究員を経て, 2013年4月より青山学院大学理工学部助教, 現在に至る。ユビキタス発電システム, 自律分散制御, ロボティックセンサネットワーク, 生体信号処理, ロボットによる自動組立, ウェアラブルデバイス, 参加型センシングの研究に従事。日本ロボット学会, 日本機械学会, 計測自動制御学会, IEEE各会員。



斉藤 裕樹 (正会員)

2001年明治大学大学院理工学研究科基礎理工学専攻情報科学系博士課程修了。博士(工学)。明治大学, 東京電機大学を経て, 2013年明治大学総合数理学部先端メディアサイエンス学科准教授, 現在に至る。センシングと情報ネットワークのソフトウェアの研究に従事。日本ソフトウェア科学会, ACM, IEEE各会員。



戸辺 義人 (正会員)

東芝, 慶應義塾大学, 東京電機大学を経て, 2012年青山学院大学理工学部情報テクノロジー学科教授。センサネットワーク, 参加型センシングの研究を進めている。IEEE, 電子情報通信学会, 計測自動制御学会, 人間情報学会各会員。